

Like what you see: Generalization of social learning determines art appreciation

Yannick Boddez^{1, 2}, Mathilde Descheemaeker^{1, 2}, Gaëtan Mertens³, Astrid Truys², Sander Van de Cruys⁴

¹Department of Clinical Psychology and Experimental Psychopathology, University of Groningen, Groningen, The Netherlands

²Centre for the Psychology of Learning and Experimental Psychopathology, KU Leuven, Leuven, Belgium

³Department of Clinical Psychology, Utrecht University, Utrecht, the Netherlands

⁴Laboratory of Experimental Psychology, Brain & Cognition Unit, KU Leuven, Belgium

Correspondence concerning this article can be addressed to Yannick Boddez, Department of Clinical Psychology and Experimental Psychopathology, Grote Kruisstraat 2/1, 9712 TS Groningen, the Netherlands.

E-mail: y.boddez@rug.nl

Tel: [0031503634752](tel:0031503634752)

Abstract

We examine whether a stimulus generalization framework can provide insight in how experience shapes evaluative responses to artworks. Participants received positive information about one artwork and negative information about another artwork. Afterwards, we tested their evaluative responses not only to these artworks but also to similar artworks, which allowed us to assess generalization. Results showed that the artwork that was paired with positive information and the artwork that was similar to it were evaluated more positively than the other artworks. These findings confirm that theories that aim to explain art appreciation could benefit from taking learning and its generalization into account.

Keywords: art; aesthetic preference; evaluative learning; generalization; evaluative priming

Like what you see: Generalization of social learning determines art appreciation

Evaluative responses determine many aspects of our daily life, including the company that we keep, the products that we buy, and the food that we eat (De Houwer, Thomas, & Baeyens, 2001). Not surprisingly, scholars have long tried to understand the origins of these responses (Martin & Levey, 1994; Zajonc, 1980). One domain in which evaluative responses can be especially puzzling is that of modern and contemporary art, which has been argued to be inaccessible and difficult (Wolfe, 1975).

We examine whether a generalization framework can provide insight in how experience shapes evaluative responses to certain artworks. Investigating the role of experience has long remained an unmet challenge in psycho-aesthetics, since the dominant strategy has been to seek for universal laws which link variations in stimulus features to variations in evaluation (Martindale, 1990; Ramachandran, 2001; Redies, 2007). Although this strategy has a long history that can be traced back to ideas about the golden ratio or divine proportion (e.g., Pacioli, 1509), most theorists now do acknowledge that experience (e.g., in the form of familiarity and training) has considerable influence (Jacobsen & Beudt, 2017; Pelowski, Markey, Forster, Gerger, & Leder, 2017; Reber, Schwarz, & Winkielman, 2004). However, exactly how the influence of experience plays out is still left unexplored. The generalization of acquired evaluation from one stimulus to other but related stimuli provides one straightforward principle of how experience might shape evaluative responding. Generalization of *negative* evaluations has previously been demonstrated in the aversive learning domain (Boddez, Bennett, van Esch, & Beckers, 2017). For instance, a dark grey square (i.e., a generalization stimulus or GS) can come to elicit dislike due to a learning experience in which a black square was followed by electric shock (i.e., a conditional stimulus or CS; Boddez et al., 2017; Dymond, Dunsmoor, Vervliet, Roche, & Hermans, 2015; Ghirlanda & Enquist, 2003; Hermans, Baeyens, & Verviet, 2013). It is as if the negative evaluation acquired by the black square spreads out to similar stimuli. However, the generalization of *positive* evaluations has rarely been investigated (FeldmanHall et al., 2018; Verosky & Todorov, 2010; Zizak & Reber, 2004) and in particular not with artworks as stimuli.

In summary, the generalization framework entails that an acquired evaluation will not remain specific to the stimulus at hand but will affect the evaluation of related stimuli as well. Interestingly, generalization allows to anchor several ideas and observations about art appreciation in a unifying framework. Below, we substantiate this claim with some examples.

First, the sometimes-observed more positive evaluation of figurative art relative to abstract art (Lindell & Mueller, 2011; Van Paaschen, Bacci, & Melcher, 2015) can be

understood as an instance of generalization in case that the figurative artwork depicts — and therefore is similar to — already liked things (e.g., an actual landscape which is lovely). In such case, transfer of the positive evaluation is indeed expected according to the generalization principle.

Second, it has been hypothesized (Loewy, 1951) that people tend to like art and design which is not too different from what they already know and like. An intuitive example is the popularity of remixes or sampling in which elements of an original song are taken and altered (example from Hekkert, 2006). More anecdotally, this can also be recognized in people liking the whole oeuvre of a certain artist or even complete art movements, since similarity within an oeuvre or movement is typically high. Both these phenomena are in line with an empirically established moderator of generalization: The strength of generalization effects is inversely related to physical distance from the original stimulus. So, the more similar a GS is to a CS, the more transfer of responding. For example, in the study discussed above, a dark grey square elicited more dislike than a light grey one after the black square was followed by electric shock (Boddez et al., 2017). Making the leap to the domain of aesthetics, a piece of art which remains closer in similarity to an already liked artwork would be predicted to elicit more liking than a less similar one.

Third and in the same vein, generalization can even shed light on the evolution of art history. It has been argued that artistic styles that differ a bit – but, importantly, not too much – from what is mainstream are typically highly appreciated and hailed as the “new thing” (Wolfe, 1975). The bull series by artist Roy Lichtenstein (*Figure 1*), which comments on this gradual evolution of mainstream styles, illustrates this idea: The realistic bull (lower right panel) has little features in common with the abstract bull (upper left panel), but the cubistic bull (lower left panel) bridges the gap between the two. Generalization allows to explain this pattern. An artistic style that is still similar to an already appreciated style will be appreciated as well. With a few intermediate steps, one can, however, end up with a style that seemingly has nothing in common with the original style.



Figure 1. Bull series by Roy Lichtenstein.

All the above still leaves open the question of where initial evaluative responses come from and how the positive evaluation of the intermediate steps is further augmented so that it can move along to increasingly different artworks. Learning theory can rely on several learning principles here, such as (further) mere exposure (e.g., through media; Zajonc, 1969) and social learning (Rachman, 1977). Social learning is especially relevant for our present purposes. Art is part of one's sociocultural surroundings and people share their reflections orally or in written reviews, which provides ample opportunity for this type of learning.

In the present study, we therefore paired artworks with positive or negative messages provided by experts. Given a wide literature on persuasive messages (De Houwer & Hughes, 2016; Kruglanski & Thompson, 1999; Wood, 2000) and a previous demonstration that such procedure can install art appreciation (Lauring et al., 2016), we figured that this would provide us with robust initial learning that would allow to assess subsequent generalization of evaluative responses. The experiment started with an evaluative learning phase, in which participants received positive information about one artwork (CS+) and negative information about another artwork (CS-). To enhance ecological validity, we used pictures of artworks that are relevant in today's art market (Hayn-Leichsenring, 2017). We hypothesized that participants would rate the artwork about which they received positive information as better than the artwork about which they received negative information. Crucially, they also had to rate a first additional artwork that was similar to the one about which they received positive information (GS+) and a second additional artwork that was similar to the one about which they received negative information (GS-). This allowed us to assess the generalization of evaluative responses. In addition to asking participants to rate the different artworks, we also

measured evaluative responding to these artworks indirectly using the Evaluative Priming Task (EPT; Fazio, Sanbonmatsu, Powell, & Kardes, 1986). This task requires participants to categorize target stimuli as positive or negative as quickly as possible. Each target is preceded by a prime stimulus for which the evaluation is under investigation (in this study, the CSs and GSs). There is evidence showing that categorization of the target is faster when the prime and the target are affectively congruent compared to when these stimuli are affectively incongruent (Fazio, 2001; Herring et al., 2013). Accordingly, we used the speed with which participants categorize positive as compared to negative targets as an indication of automatic evaluative responding towards the CSs and GSs.

Method

Participants

Twenty-four participants (18 women), between 19 and 60 years old ($M = 28.58$ years, $SD = 12.46$ years), were recruited through an online system that allows both people from the community and students to sign up for experiments. The incentive for participating was either a financial reimbursement of 10 euro or a course credit. All participants gave written informed consent. Sample size was determined in advance based on within-group comparisons in previous evaluative conditioning studies (e.g., Boddez et al., 2017). Participants were assigned to one of the two counterbalancing groups in an alternating manner. The protocol was approved by the social and societal ethics committee at KU Leuven and carried out in accordance with the provisions of the World Medical Association Declaration of Helsinki.

Apparatus and Stimuli

Participants were tested in individual sessions on a Dell desktop computer with a CRT monitor. Affect 4.0 software (Spruyt, Clarysse, Vansteenwegen, Baeyens, & Hermans, 2010) was used to control stimulus presentation and register responses.

In the rating phase, we used eight artworks, with dimensions varying between 297 by 600 pixels and 512 by 600 pixels. In counterbalancing condition 1, the CS+ was an artwork by Dirk Braeckman, named *M.R.-P.O.* (top panel of Figure 2), and the CS- was an artwork by Francis Bacon, named *Study for head of Lucian Freud* (bottom panel of Figure 2). In counterbalancing condition 2, this was reversed. The GS for the artwork by Dirk Braeckman was an artwork by Daisuke Yokota, named *Untitled* (top panel of Figure 2), and the GS for the artwork by Francis Bacon was an artwork by Adrian Ghenie, named *Selfportrait as Vincent Van Gogh 2* (bottom panel of Figure 2). The remaining four artworks functioned as filler stimuli and were by the artists Michaël Borremans (*The prop*), Marlene Dumas (*The painter*), Geert Goiris (*Dead bird*) and Wolfgang Tillmans (*Dürerstrasse*).

To account for pre-existing individual differences, participants rated a series of artworks before they came to the experiment. This pre-rating survey was administered online with the help of LimeSurvey service and included the same eight stimuli used in the experiment and 10 additional filler artworks. These additional artworks varied between 263 by 532 pixels and 399 by 600 pixels and were by the following artists: Fred Bervoets (*Hommage aan Van Gogh*), Ellen De Meutter (*Selfportrait 2*), Marlene Dumas (*Destino*), Karin Hanssen (*The approach*), Gideon Kiefer (*You got up the tree, so you can get down*), Sophie Kuijken (*Untitled*), Hervé Martijn (*The distance*), Laurent Millet (*Somnium*), Gerhard Richter (*Frau mit Hund am See*), and Mircea Suciú (*Still life*).

During the EPT, a response box with two keys was used to categorize the targets. Targets were six positive pictures (i.e., baby, dog, cat, waterfall, dolphin, sunset) and six negative pictures (i.e., car-wreck, fire, skull, worms, barbed wire, trash) with a dimension of 512 by 384 pixels. These targets were selected on the basis of norm data collected in previous research (Spruyt, Hermans, De Houwer, & Eelen, 2002). Primes consisted of the two CSs, the two GSs, and one neutral picture (i.e., a triangle on a gray background). During the practice trials, primes were three geometric figures on a gray background (i.e., circle, square, rhombus).



Figure 2. The artwork on the top left is by Dirk Braeckman (Courtesy Zeno-X gallery, Antwerp) and was used as a CS+ in counterbalancing condition 1 and as a CS- in counterbalancing condition 2. The artwork on the top right is by Daisuke Yokota (Courtesy Stieglitz 19, Antwerp) and was used as its GS in both counterbalancing conditions. The artwork on the bottom left is by Francis Bacon and was used as CS- in counterbalancing condition 1 and as CS+ in counterbalancing condition 2. The artwork on the right is by Adrian Ghenie (Courtesy Tim Van Laere Gallery, Antwerp) and was used as its GS in both counterbalancing conditions.

Measures

(Pre-)rating. During the online survey that was completed at home before the start of the experiment and during the rating phase of the experiment, participants were asked to evaluate the artworks on a scale ranging from 0% (*very bad*) to 100% (*very good*).

Evaluative Priming Task (EPT). The EPT measured the reaction time to categorize the target pictures as positive or negative.

Similarity. Participants rated the similarity of pairs of stimuli used on a scale ranging from 0 (*not similar at all*) to 10 (*very similar*). More precisely, they were asked how alike the artworks were.

Expertise. To assess their level of expertise, participants were asked to answer two questions about the artworks they had seen during the rating phase. The first question concerned whether they had already seen the artwork before participating in the study. The participant could answer this with ‘yes’, ‘no’ or ‘not sure’. The second question asked whether they knew the artist who made the artwork, which they could answer with ‘no’ or the name of the artist. An expertise score was determined by giving one point for every artwork they had seen before, 0.5 point for every artwork of which they were not sure if they had seen it before, and one point for every correct artist name.

Procedure

Table 1 provides an overview of essential elements of the procedure. Details are discussed below.

Table 1

Essential elements of the procedure

Pre-rating (<i>very bad to very good</i>)	Evaluative learning	Rating (<i>very bad to very good</i>)	Evaluative priming task	Similarity ratings	Expertise (<i>seen before? / name artist?</i>)
CS+ CS- GS+ GS-	CS+ & positive information CS- & negative information	CS+ CS- GS+ GS-	CS+ CS- GS+ GS-	CS*GS combinations	CS+ CS- GS+ GS-

Note. Participants received positive information about one artwork (CS+) and negative information about another artwork (CS-) in the evaluative learning phase. Afterwards, we tested evaluative responding to these artworks and to two artworks similar to them (respectively GS+ and GS-) by means of a rating scale and an evaluative priming task. With respect to the rating data, we calculated the difference between post-ratings and pre-ratings to account for pre-existing differences. At the end of the experiment, participants rated the similarity of each CS with the two GSs and were tested for their expertise in art.

After signing up for the study, participants received an e-mail with a link to the online pre-rating survey in which they were asked to rate artworks for the first time. These artworks were the eight artworks that were also presented during the rating phase and the 10 extra filler artworks described above. Participants were instructed to fill out this survey between 12 to 24 hours before the start of the actual experiment. The stimuli were presented in a random order.

The experiment started with the evaluative learning training. Participants were told that they would have to rate artworks and that, as an example, they would first view two artworks accompanied by the evaluation given by experts as found on Artnet (a metacritic website about art). In fact, these example trials were the crucial learning trials. Participants were shown four trials in a random order: two times the artwork by Francis Bacon and two times the artwork by Dirk Braeckman (see Figure 2). On each trial, the artworks were accompanied by the mean rating and reviews of either an expert group of art dealers or an expert group of other artists. One of the two artworks was given a positive mean rating (i.e., 87% and 93%) and review by both expert groups, while the other artwork received a negative mean rating (i.e., 8% and 12%) and review by both expert groups. Each review consisted of three positive words / statements (i.e., intense, suggestive image, good complexity, timeless, intriguing, strong imagery) or three negative words / statements (i.e., grotesque, boring, banal portrait, stylistically outmoded, too commercial, something you would put on a postcard). The manipulation was counterbalanced: In the first counterbalancing condition, the artwork by Braeckman received a high rating by both expert groups (i.e., the CS+) and the artwork by Bacon received low ratings (i.e., the CS-), while the inverse was true in the second counterbalancing condition. Each artwork appeared for 30 seconds together with the rating and reviews placed underneath. To enhance the effect of the manipulation, the ratings and reviews were read aloud by the experimenter.

Participants were then asked to rate different artworks themselves, including the CSs (artwork by Dirk Braeckman and by Francis Bacon), the GSs (artwork by Daisuke Yokota and by Adrian Ghenie) and the four filler stimuli described above. Artworks appeared on the screen one by one in a random order. Participants indicated their rating on the scale and pressed the space bar to proceed to the next trial.

Subsequently, the EPT was administered. Each trial started with a 500-ms presentation of a fixation cross followed by a 500-ms blank interval, after which the prime was presented for 200 ms. The target was presented 50 ms after the offset of the prime. Participants were instructed to look at the prime and to categorize the target as quickly as possible as positive or negative by pressing respectively the right or left key of the response box. A trial ended when a response key had been pressed or after 2000 ms had elapsed. The inter-trial interval varied semi-randomly between 500 and 1500 ms, with an average set to 1000 ms. The task started with 12 practice trials during which each target was presented once in combination with one of the three geometric figures serving as primes. The actual priming trials were divided into two blocks separated by a reminder to continue categorizing the targets as quickly as possible. In each block, each of the five primes (i.e., CS+, CS-, GS+, GS-, geometric figure) was

combined with each of the twelve targets (i.e., six positive pictures and six negative pictures), resulting in 60 trials per block that were presented in a random order. After this task, participants rated the targets on a scale ranging from -10 (*very negative*) to 10 (*very positive*) in order to verify their intended positive or negative evaluation.

During the next phase, participants rated the similarity of each CS with the two GSs and the four filler stimuli. Artworks were presented in pairs on the computer screen in a random order. Participants rated their similarity on the scale and pressed the space bar to proceed to the next trial. A successful manipulation would point out that (a) the artworks by Adrian Ghenie and Francis Bacon look more similar to each other than the artworks by Adrian Ghenie and Dirk Braeckman, and that (b) the artworks by Daisuke Yokota and Dirk Braeckman look more similar to each other than the artworks by Daisuke Yokota and Francis Bacon.

Finally, to assess participants' pre-existing expertise in art, they were instructed to indicate whether they had already seen the eight artworks of the rating phase before the onset of the experiment and to write down the names of the corresponding artists. The artworks were shown on the screen and participants wrote down their answers on a piece of paper they received.

Results

Ratings

To account for pre-existing individual differences in evaluative responding, we decided before the start of data collection to calculate the difference between post-ratings and pre-ratings for the CS+, the CS-, the GS+, and the GS-. We then performed a repeated measures ANOVA on these difference scores with two within-subjects variables: intended evaluation (positive for the CS+ and the GS+ versus negative for the CS- and GS-) and stimulus type (CS versus GS). Figure 3 shows that the CS+ and the GS+ were rated higher than the CS- and the GS-, as hypothesized. The analysis confirmed the presence of a significant effect of intended evaluation, $F(1, 23) = 5.98$, $MSE = 360.42$, $\eta^2_p = .21$, $p = .023$. This effect of intended evaluation did not differ between CSs and GSs, as indicated by the absence of an interaction effect between stimulus type and intended evaluation, $F(1, 23) = 0.03$, $MSE = 202.73$, $\eta^2_p = .00$, $p = .859$. There was no significant difference in ratings between the CSs and the GSs, as indicated by the absence of an effect of stimulus type, $F(1, 23) = 2.82$, $MSE = 4688.99$, $\eta^2_p = .11$, $p = .107$.

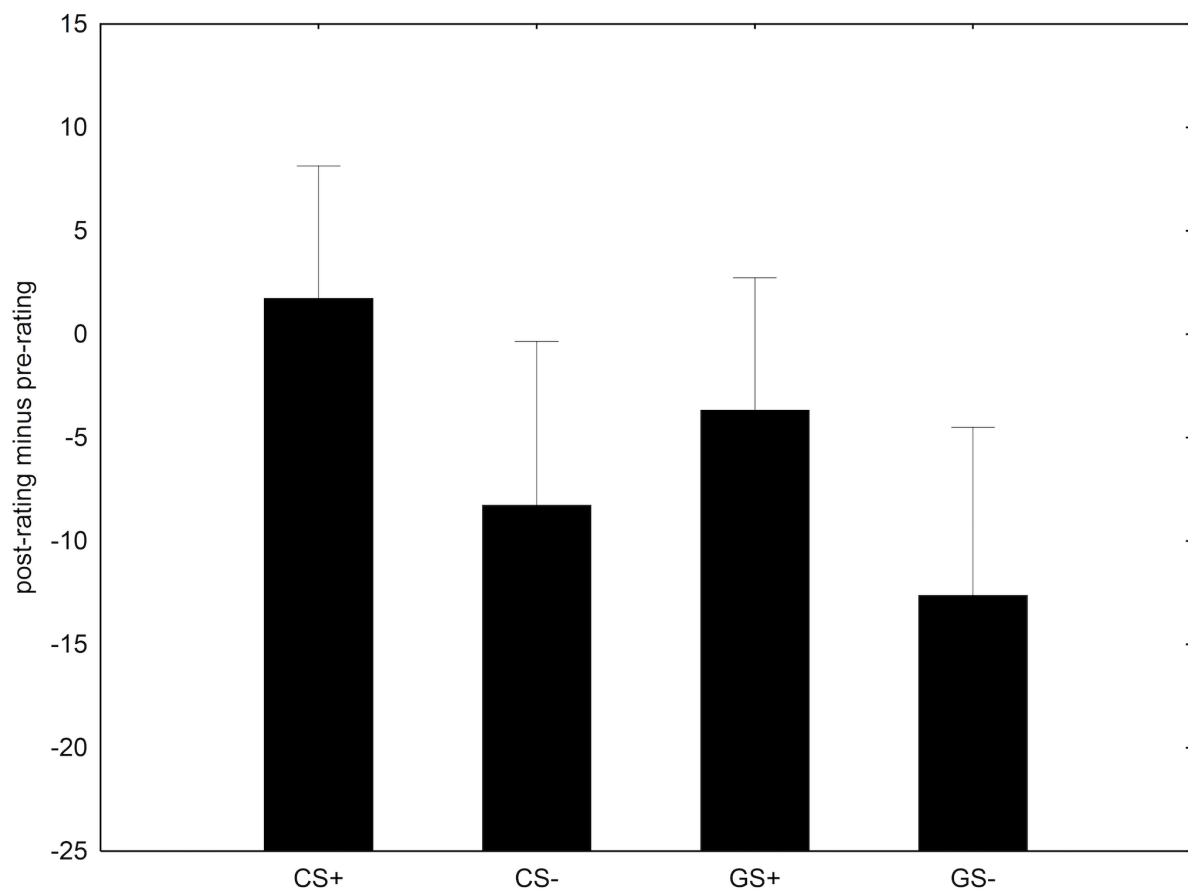


Figure 3. Mean difference scores for the CS+, CS-, GS+ and GS- (see main text for details). Error bars represent 95% confidence intervals.

Evaluative Priming Task

The method of data pre-processing and all exclusion criteria were determined beforehand and based on Descheemaeker, Spruyt, and Hermans (2014). The manipulation check at the end of the EPT indicated that participants rated the positive targets as significantly more positive than the negative targets, $M_{\text{positive}} = 7.5$, $SD_{\text{positive}} = 0.4$, $M_{\text{negative}} = -7.5$, $SD_{\text{negative}} = 1.2$, Welch t-test: $t(5.9) = 28.35$, $p < .001$. However, one participant rated a negative target as neutral, so all trials with this target were excluded for this participant. We also excluded trials on which a target was categorized incorrectly (3.2% of all trials), trials on which no response was given before 2000 ms had elapsed (0.2% of all trials), and trials with a reaction time of 100 ms or less (0.1% of all trials). To reduce the impact of outlying variables, we excluded response latencies that deviated more than 2.5 standard deviations from a participant's mean latency for each trial type (e.g., trials with the CS+ as a prime followed by a positive target; 2.5% of all

trials). Based on the remaining trials, we calculated mean response latencies for each trial type per participant.

These mean response latencies were analyzed by means of a 2 (prime type: GS versus CS) x 2 (intended prime evaluation: positive for the CS+ and the GS+ versus negative for the CS- and GS-) x 2 (target evaluation: positive versus negative) repeated measures ANOVA. A test of the interaction between intended prime evaluation and target evaluation resulted in $F(1,23) = 3.90$, $MSE = 1199.97$, $\eta^2_p = .15$, $p = .061$. Although this p-value is just above the commonly employed significance level of 5 per cent, the data pattern was partly in line with our hypothesis when considered at a mere descriptive level. As can be seen in Figure 4, participants categorized positive targets more quickly than negative targets when they had first been presented with the CS+ or the GS+. In the case that the CS- or GS- had been presented first, this difference in speed of responding was less pronounced. However, in this case, we did not see a full reversal of the pattern of responding as compared to when the CS+ or GS+ had been presented first though (i.e., participants were not faster to categorize negative as compared to positive targets). This lack of a full reversal might be because priming effects are generally stronger for positive stimuli (Unkelbach, Fiedler, Bayer, Stegmüller, & Danner, 2008) or merely because EPT scores tend to be relatively low in reliability (Van Dessel, Mertens, Smith, & De Houwer, 2017).

The absence of a significant three-way interaction effect between prime type, intended prime evaluation, and target evaluation, $F(1,23) = 0.06$, $MSE = 909.13$, $\eta^2_p = .00$, $p = .803$, indicates that the interaction between intended prime and target evaluation was not different for CSs as compared to GSs. Needless to say, this three-way interaction needs to be interpreted with caution given that the p-value of the interaction between intended prime and target evaluation was above the commonly employed significance level.

For the sake of completeness, we still report the main effect of prime type (i.e., CS versus GS), $F(1,23) = .01$, $MSE = 5$, $\eta^2_p = .00$, $p = .93$, of intended prime evaluation (i.e., CS+ and GS+ versus CS- and GS-), $F(1,23) = .07$, $MSE = 213$, $\eta^2_p = .00$, $p = .80$, and of target evaluation (i.e., positive versus negative targets), $F(1,23) = 9.21$, $MSE = 22639$, $\eta^2_p = .90$, $p = 0.01$.

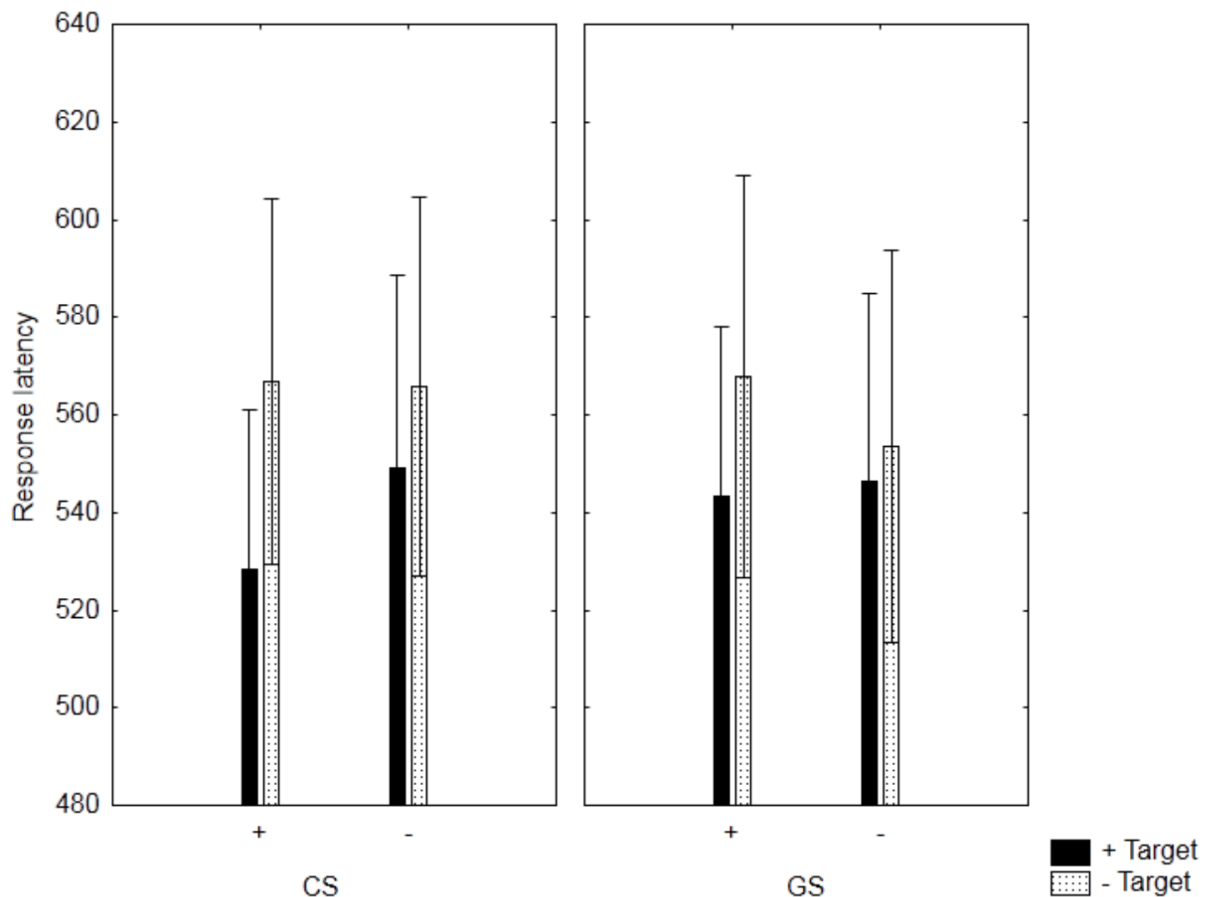


Figure 4. Mean response latencies in ms during the Evaluative Priming Task (see main text for details). Error bars represent 95% confidence intervals.

Similarity

Similarity ratings of each CS with each GS were analyzed by means of a repeated measures ANOVA with one within-subjects variable (similarity: similarity between Francis Bacon and Daisuke Yokota, between Dirk Braeckman and Adrian Ghenie, between Francis Bacon and Adrian Ghenie, and between Dirk Braeckman and Daisuke Yokota). Figure 5 shows that the artworks that were supposed to look similar were actually rated as more similar. The results indeed revealed a significant effect of similarity, $F(2.12, 48.75) = 49.96$, $MSE = 241.83$, $\eta^2_p = .68$, $p < 0.001$. Follow-up tests confirmed that the artworks by Francis Bacon and Adrian Ghenie looked more similar than the ones by Dirk Braeckman and Adrian Ghenie, $F(1,23) = 126.23$, $MSE = 3.52$, $\eta^2_p = .85$, $p < 0.001$, and that the artworks by Dirk Braeckman and Daisuke Yokota looked more similar than the ones by Francis Bacon and Daisuke Yokota, $F(1,23) = 30.79$, $MSE = 8.79$, $\eta^2_p = .57$, $p < .001$.

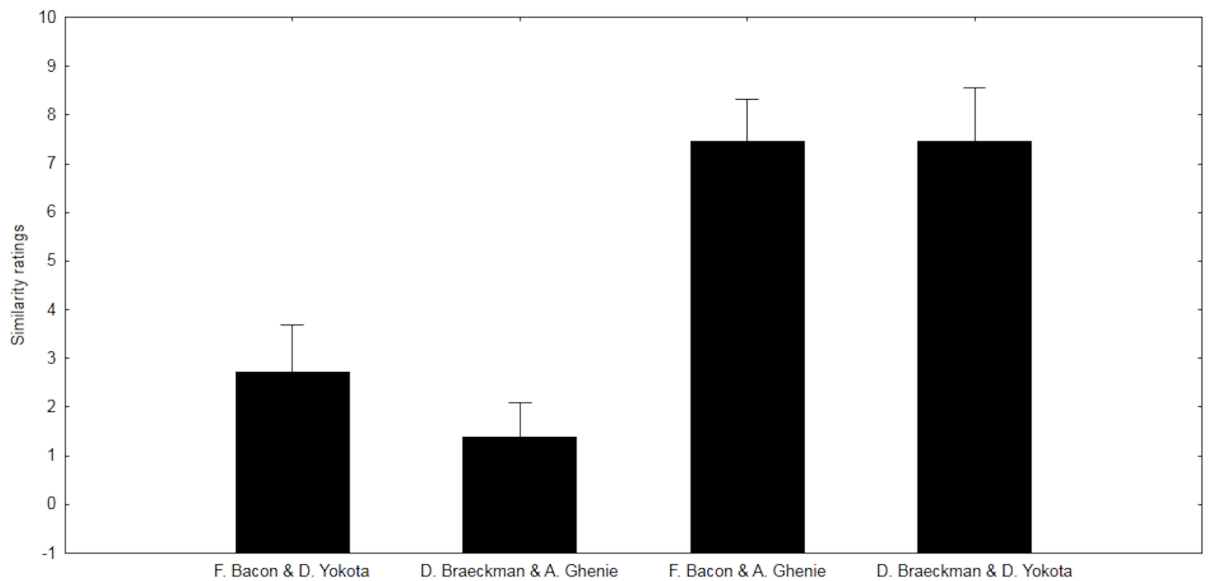


Figure 5. Mean similarity ratings between each CS and GS. Error bars represent 95% confidence intervals.

Expertise

Descriptive statistics for the expertise scores were $M = 0.65$ and $SD = 1.15$. The highest expertise score was 2.5 (this participant indicated having seen one of the artworks before the start of the experiment and was not sure if he/she had seen three of the other artworks before). None of the participants could state the name of an artist who created one of the artworks used during the experiment. We can conclude that there were no experts among our participants.

Discussion

We hypothesized that evaluative responses to artworks could be installed by providing verbal information about these artworks. Crucially, we hypothesized that this learning effect would also generalize to other, similar artworks. The rating data confirmed our hypotheses. Participants rated the CS+ and the GS+ as better than the CS- and GS-. At a descriptive level, the pattern of the EPT data corroborated this finding, as the time needed to categorize a target was moderated by whether either a CS+ or a GS+, on the one hand, or a CS- or a GS-, on the other hand, preceded it. However, the p-value of the corresponding interaction remained just above the commonly used significance level of 5 per cent.

In line with the assumptions of the functional-cognitive framework (e.g., De Houwer, Gawronski, & Barnes-Holmes, 2013), we have to differentiate between the functional and the cognitive level of analysis when interpreting these results. At the functional level, we found evidence that information about artworks affects the evaluative responses to these and related

artworks. At the cognitive level, this effect could be mediated by different candidate mechanisms such as the formation of associations between the positive/negative words in the reviews (e.g., “powerful”, “good”, and “intriguing” versus “boring”, “banal”, and “outmoded”) and the artwork (for such a proposal see Gast & De Houwer, 2012), inferences made by the participants regarding the plausibility and trustworthiness of the expert reviews (De Houwer, 2018), or even with experimental demand. With respect to experimental demand, two arguments are worth considering. First, it is of note that the EPT which we used is considered to be insensitive to control strategies that could lead to demand effects (Gawronski, Mitchell, & Balas, 2015). The convincing power of this first argument will of course depend on how convincing one finds the data pattern obtained with this task. Second, effects of experimental demand presuppose the existence of a straightforward desired response. Arguably, the present research design involves a certain degree of ambiguity: Participants might infer which response is expected to the CSs, but it might not be so straightforward to infer which response is expected to the GSs.

The analyses for the rating data were performed on difference scores between post-ratings and pre-ratings. The (on average) negative difference scores suggest that participants tended to give higher ratings in the pre-rating phase (which was completed at home before the start of the experiment) than during the rating phase of the experiment. We can only speculate on why this is so. It could have to do with the context (home versus in the laboratory) or with rating the artworks for the first versus for the second time. Some could consider the latter an unlikely candidate-explanation, because repeated exposure has often been linked to more positive evaluations (mere exposure and fluency effects; Berlyne, 1970; Van de Cruys, & Wagemans, 2011; Zajonc, 1969).

The generalization findings suggest that the impact of an evaluative learning experience can spread out to other stimuli: Positive or negative information about one artwork can affect one’s evaluation of similar artworks. This is in line with conditioning models that predict generalization to stimuli that share perceptual features with the original CS (e.g., Pearce, 1987). These models assume that a stimulus is composed of multiple perceptual elements that can each, to a certain degree, be associated with an outcome (e.g., an electric shock or a positive message). Accordingly, GSs can activate the outcome representation to the extent that they are composed of elements that are also part of the CS (e.g., Pearce, 1987; Rescorla, 1976; but see Boddez et al., 2017). It is of note that in the present study the generalization effect may be driven by perceptual features concerning style (e.g., color palette) and / or content (e.g., shape of a head versus of a body).

It is hereby important to note that the current design cannot ascertain that similarity between artworks is driving the congruent evaluation of CSs and GSs. The similarity ratings at the end of experiment do suggest this but provide no causal evidence. However, it is not immediately clear what the common ground between the CSs and GSs that were evaluated congruently would be if not similarity. In addition, given the broader literature on generalization, we would argue that it seems more likely that similarity drives evaluative responses as compared to, for example, the other way around. In generalization studies with stimuli that are initially neutral and arbitrary (so not artworks), physical distance (and thereby presumably similarity) is manipulated by testing stimuli that vary systematically over some stimulus dimension (e.g., intensity of sound/light, orientation, object size, etc.; Ghirlanda & Enquist, 2003). Evaluative responses to these stimuli have been observed to vary as a function of distance of the original CS (Boddez et al., 2017), suggesting that similarity causes congruent evaluation (of initially neutral stimuli), rather than the other way around.

Although it is possible that the current results are specific to the used stimulus materials, we consider this unlikely given that generalization has proven to be a robust phenomenon in several domains other than art appreciation (Dymond et al., 2015; Ghirlanda & Enquist, 2003). It would nonetheless be advised to expand the stimulus set in future research. Adding more GSs would also allow to assess whether there would be a decline in positive evaluation of artworks that become increasingly more dissimilar. The strength of generalization effects is indeed known to be inversely related to physical distance from the original stimulus (Dymond et al., 2015; Ghirlanda & Enquist, 2003).

Another suggestion for future research is to assess generalization across non-perceptual dimensions. In the present study, the CS and the GS were perceptually similar. However, in daily life, art appreciation also seems to spread across non-perceptual dimensions. People might, for example, like all artworks from a certain time period or all artworks sold by a certain prestigious art dealer (Wolfe, 1975), although there might be little perceptual overlap between these artworks. Different procedures from the learning tradition are highly relevant for understanding and studying this type of generalization. These procedures offer a venue for an induction of equivalent classes of stimuli, that is, sets of completely arbitrary stimuli, bearing no perceptual relation, that nevertheless may support within-class generalization (Hermans & Baeyens, 2013; Hughes, De Houwer, & Perugini, 2016). Future research could make use of these procedures to further our understanding of generalization of art appreciation along conceptual similarity dimensions as well.

A phenomenon related to generalization is peak shift. Simply put, the peak shift effect concerns a shift of peak responding away from the CS+ in the direction opposite of the CS–.

For example, in a fear conditioning procedure, if a small circle is followed by a shock and a big circle is not followed by a shock, an even smaller circle will provoke more fear than the small circle that was actually followed by the shock (Struyf, Iberico, & Vervliet, 2014). In art, one could assume, for example, that if abstract art is made positive and a figurative art is made negative (e.g, because of a trend in reviewing among art critics), even more abstract art will be evaluated as even more positive. Although this principle is considered to be important in explaining evaluative responses to specific artworks (Ramachandran & Hirstein, 1999), it has not yet been investigated in this context (Leder et al., 2004). Since the occurrence of peak shift presupposes the occurrence of generalization, our study can be seen as an incentive to look into peak shift effects in follow-up studies on art appreciation.

In summary, we demonstrated that generalization can help us understand how experience shapes evaluative responses to artworks. Instructed art appreciation spread out to similar but different artworks.

Acknowledgements

The authors are grateful to Tom Beckers and Jan De Houwer for helpful discussions about the topics at hand. Data was collected at KU Leuven and supported by KU Leuven Program Funding Grant PF/10/005. Sander Van de Cruys is a postdoctoral fellow of the Research Foundation – Flanders (FWO).

References

- Boddez, Y., Bennett, M., P., van Esch, S., & Beckers, T. (2017). Bending rules: The shape of the perceptual generalization gradient is sensitive to inference rules. *Cognition & Emotion*, 31, 1444-1452.
- Berlyne, D. E. (1970). Novelty, complexity, and hedonic value. *Perception & Psychophysics*, 8, 279-286.
- De Houwer, J. (2018). Propositional models of evaluative conditioning. *Social Psychological Bulletin*, 13, e28046.
- De Houwer, J., Gawronski, B., & Barnes-Holmes, D. (2013). A functional-cognitive framework for attitude research. *European Review of Social Psychology*, 24, 252-287.
- De Houwer, J., & Hughes, S. (2016). Evaluative conditioning as a symbolic phenomenon: On the relation between evaluative conditioning, evaluative conditioning via instructions, and persuasion. *Social Cognition*, 34, 480-494.
- De Houwer, J., Thomas, S., & Baeyens, F. (2001). Association learning of likes and dislikes: A review of 25 years of research on human evaluative conditioning. *Psychological Bulletin*, 127, 853-869.
- Descheemaeker, M., Spruyt, A., & Hermans, D. (2014). On the relationship between the indirectly measured attitude towards beer and beer consumption: The role of attitude accessibility. *PloS ONE*, 9, 1-8.
- Dymond, S., Dunsmoor, J. E., Vervliet, B., Roche, B., & Hermans, D. (2015). Fear generalization in humans: systematic review and implications for anxiety disorder research. *Behavior Therapy*, 46, 561-582.
- Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology*, 50, 229-238.
- Fazio, R. H. (2001). On the automatic activation of associated evaluations: An overview. *Cognition & Emotion*, 15, 115-141.
- FeldmanHall, O., Dunsmoor, J. E., Tompary, A., Hunter, L. E., Todorov, A., & Phelps, E. A. (2018). Stimulus generalization as a mechanism for learning to trust. *Proceedings of the National Academy of Sciences*, 115, E1690-E1697.
- Gast, A., & De Houwer, J. (2012). Evaluative conditioning without directly experienced pairings of the conditioned and the unconditioned stimuli. *The Quarterly Journal of Experimental Psychology*, 65, 1657-1674.

- Gawronski, B., Mitchell, D. G., & Balas, R. (2015). Is evaluative conditioning really uncontrollable? A comparative test of three emotion-focused strategies to prevent the acquisition of conditioned preferences. *Emotion, 15*, 556 - 568.
- Ghirlanda, S., & Enquist, M. (2003). A century of generalization. *Animal Behaviour, 66*, 15-36.
- Hayn-Leichsenring, G. U. (2017). The ambiguity of artworks - A guideline for empirical aesthetics research with artworks as stimuli. *Frontiers in Psychology, 8*, 1-14.
- Hekkert, P. (2006). Design aesthetics: principles of pleasure in design. *Psychology science, 48*, 157-172.
- Hermans, D., & Baeyens, F. (2013). Generalization as a basis for emotional change: Perceptual and non-perceptual processes. In D. Hermans, B. Rimé, & B. Mesquita (Eds.), *Changing Emotions* (pp. 67-73). Hove, UK: Psychology Press.
- Hermans, D., Baeyens, F., & Vervliet, B. (2013). Generalization of Acquired Emotional Responses. In M. D. Robinson, E. R. Watkins, & E. Harmon-Jones, *Handbook of Cognition and Emotion* (pp. 117 - 132). New York: Guilford Press.
- Herring, D. R., White, K. R., Jabeen, L. N., Hinojos, M., Terrazas, G., Reyes, S. M., ... & Crites Jr, S. L. (2013). On the automatic activation of attitudes: A quarter century of evaluative priming research. *Psychological Bulletin, 139*, 1062-1089.
- Hughes, S., De Houwer, J., & Perugini, M. (2016). Expanding the boundaries of evaluative learning research: How intersecting regularities shape our likes and dislikes. *Journal of Experimental Psychology: General, 145*, 731-754.
- Jacobsen, T., & Beudt, S. (2017). Stability and Variability in Aesthetic Experience: A Review. *Frontiers in psychology, 8*, 1-14.
- Kruglanski, A. W., & Thompson, E. P. (1999). Persuasion by a single route: A view from the unimodel. *Psychological Inquiry, 10*, 83-109.
- Lauring, J. O., Pelowski, M., Forster, M., Gondan, M., Ptito, M., & Kupers, R. (2016). Well, if they like it... Effects of social groups' ratings and price information on the appreciation of art. *Psychology of Aesthetics, Creativity, and the Arts, 10*, 344.
- Leder, H., Belke, B., Oeberst, A., & Augustin, D. (2004). A model of aesthetic appreciation and aesthetic judgments. *British Journal of Psychology, 95*, 489-508.
- Lindell, A., & Mueller, J. (2011). Can science account for taste? Psychological insights into art appreciation. *Journal of Cognitive Psychology, 23*, 453-475.
- Loewy, R. (1951). *Never Leave Well Enough Alone*. New York: Simon and Schuster.
- Martin, I., & Levey, A. (1994). The evaluative response: Primitive but necessary. *Behaviour Research and Therapy, 32*, 301-305.

- Martindale, C. (1990). *The Clockwork Muse: The Predictability of Artistic Change*. New York: Basic Books.
- Pacioli, L. (1509). *The Divine Proportion*. Venice.
- Pearce, J. M. (1987). A model for stimulus generalization in Pavlovian conditioning. *Psychological Review*, 94, 61-73.
- Pelowski, M., Markey, P. S., Forster, M., Gerger, G., & Leder, H. (2017). Move me, astonish me... delight my eyes and brain: The Vienna integrated model of top-down and bottom-up processes in art perception (VIMAP) and corresponding affective, evaluative, and neurophysiological correlates. *Physics of Life Reviews*, 21, 80-125.
- Rachman, S. (1977). The conditioning theory of fear acquisition: A critical examination. *Behaviour Research and Therapy*, 15, 375–387.
- Ramachandran, V. S. (2001). Sharpening up ‘The science of art’. *Journal of Consciousness Studies* 8, 9–29.
- Ramachandran, V. S., & Hirstein, W. (1999). The science of art: A neurological theory of aesthetic experience. *Journal of Consciousness Studies*, 6, 15-51.
- Reber, R., Schwarz, N., & Winkielman, P. (2004). Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience? *Personality and social psychology review*, 8, 364-382.
- Redies, C. (2007). A universal model of esthetic perception based on the sensory coding of natural stimuli. *Spatial Vision*, 21, 97-117.
- Rescorla, R. A. (1976). Stimulus generalization: Some predictions from a model of Pavlovian conditioning. *Journal of Experimental Psychology: Animal Behavior Processes*, 2, 88–96.
- Spruyt, A., Hermans, D., De Houwer, J., & Eelen, P. (2002). On the nature of the affective priming effect: Affective priming of naming responses. *Social Cognition*, 20, 227-256.
- Spruyt, A., Clarysse, J., Vansteenwegen, D., Baeyens, F., & Hermans, D. (2009). Affect 4.0. *Experimental Psychology*, 57, 36–45.
- Struyf, D., Iberico, C., & Vervliet, B. (2014). Increasing predictive estimations without further learning. *Experimental Psychology*, 61, 131-141.
- Unkelbach, C., Fiedler, K., Bayer, M., Stegmüller, M., & Danner, D. (2008). Why positive information is processed faster: the density hypothesis. *Journal of Personality and Social Psychology*, 95, 36-49.
- Van de Cruys S., Wagemans J. (2011). Putting reward in art: A tentative prediction error account of visual art. *i-Perception*, 2, 1035-1062.

- Van Dessel, P., Mertens, G., Smith, C. T., & De, J. H. (2017). The Mere Exposure Instruction Effect. *Experimental psychology*, 64, 299-314.
- Van Paasschen, J., Bacci, F., & Melcher, D. P. (2015). The influence of art expertise and training on emotion and preference ratings for representational and abstract artworks. *PloS one*, 10, 1-21.
- Verosky, S. C., & Todorov, A. (2010). Generalization of affective learning about faces to perceptually similar faces. *Psychological Science*, 21, 779-785.
- Wolfe, T. (1975). *The Painted Word*. New York: Farrar, Straus and Giroux.
- Wood, W. (2000). Attitude change: Persuasion and social influence. *Annual Review of Psychology*, 51, 539-570.
- Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist*, 35, 151-175.
- Zizak, D. M., & Reber, A. S. (2004). Implicit preferences: The role (s) of familiarity in the structural mere exposure effect. *Consciousness and Cognition*, 13, 336-362.